

# EXPERIMENTS ON THE RATIONALE OF THE SZONDI TEST

By

**PHILIP L. SHORT, M.A.**

*The Burden Neurological Institute, Bristol*

(Received 1 November, 1953)

## INTRODUCTION

THE purpose of this paper is to investigate one of the basic assumptions of the Szondi test, namely, that in selecting pictures in the test subjects tend to assign consistent meanings to the pictures, although these meanings are not explicitly stated. In the following experiments, this assumption was tested by obtaining *explicit* reactions to the pictures in the form of identifications, using two different experimental methods. Similar studies have been made by Klopfer and Borstelmann (1950), Rabin (1950a), Fosberg (1951), Davis and Raimy (1952), and by Dudek and Patterson (1952). In all these investigations, American subjects with some experience of psychology or psychiatry were used. The most extensive work was done by Klopfer and Borstelmann, who used a sample of American psychology students, and suggested that their methods might be employed on "other segments of the population". In the following experiments, their methods have been largely adopted for English subjects, of varying ages and occupations, including students.

It was realized at the outset that if it was found experimentally that consistent identifications were found for any or all of the Szondi pictures, such a finding would still throw no light on the mechanisms of choice involved in the ordinary administration of the test, as formulated by Szondi (1952) and Deri (1949). Deri, the main exponent of Szondi's theory and practice, finds it unnecessary to include Szondi's own genetic assumptions in her account of the rationale of the test. Instead, she hypothesizes that people have latent tendencies towards certain mental diseases, which show themselves in certain personality characteristics; these tendencies are responsible for the choice of particular pictures.

It is difficult to conceive of workable experiments that would either affirm or deny Szondi's or Deri's suppositions about the factors determining choices. Wallen (1951), made an initial attempt to tackle the problem experimentally by analysing one possible factor involved in choosing, namely that of facial clues and characteristics. This factor may be of some importance in explaining the results to be described, although it has not been examined specifically. The main value of the present investigation lies elsewhere, and is twofold.

Firstly, it suggests that the extent to which correct identifications are found for the pictures indicates the relative degrees of assurance a tester might adopt in scoring particular pictures or groups of pictures. Secondly, as Rabin (1950a), pointed out, such a study indicates efficiency of judgment in the identification of diagnoses based entirely on photographs, divorced from all the verbal and kinaesthetic clues provided by "live" patients. The first outcome should have important bearings on the validity of the test per se, and the second may suggest interesting implications for constructing procedures for the identification of pathological states, and hence may be of value to clinical psychologists.

## METHODS AND PROCEDURES

Two different methods and procedures were adopted to discover the way subjects interpret the Szondi pictures. These are named (a) the Description Method, and (b) the Matching Method.

*(a) Description Method*

The cards were allotted numbers from 1 to 48. Each subject was given a complete set of pictures and asked simply to write down what sort of person he thought was represented in each photograph. No limit was set to the length of the description or to the time taken to complete the experiment. No further attempt was made to explain the nature of the experiment, as it was regarded as a self-evident puzzle, wherein the "true" personality was the correct answer. The experiment was completed in one to four sittings with each subject. Occasionally, several subjects were engaged on the experiment at one and the same time, so that the order in which the sets were presented varied from one subject to another.

This method is similar to the "free association" method used by Klopfer and Borstelmann (1950), except that an attempt was made to control subjects' responses to the extent of discouraging irrelevant questions and comments from subjects, in order to facilitate scoring. It should be noted that Szondi himself used an association method to discover the stimulus value of his pictures. But he failed to specify the number and kind of subjects who provided the associations.

The "panel" rating of responses, used by Klopfer, was not possible in this experiment. Instead, a single judge was used who classified each description to fit the rationale of one of the eight Szondi factors. This did not proceed by the progressive elimination of factors, set by set. The order of presentation was randomized so that each description received a separate and distinct evaluation. If a description failed to suggest any one of the eight factors, no evaluation was made. A total of 943 out of a possible 1,029 descriptions were scored as suggesting one or other of the eight factors, leaving 86 unclassified responses.

*(b) Matching Method*

This procedure was devised as a means of countering the subjective judgments involved in the previous method, and for the purposes of easier quantification. Three separate tests were given to three groups of subjects. A checklist was cyclostyled which contained eight personality descriptions to match the eight Szondi factors. These descriptions were substantially the same as those worked out by Klopfer. They were modified only so as to replace certain obvious Americanisms by their colloquial English equivalents. The checklist had blank spaces opposite each personality description, and subjects were told that they would be shown 48 photographs, each one of which had been given a number from 1 to 48. They were to look at each picture for 30 seconds, and then place it in one of the eight categories, according to their impression as to the sort of person he or she was. Five minutes was allotted initially for subjects to read over the descriptions. The pictures were then projected episcopically on to a screen. The final result was that each subject had 48 numbers written on his form, and the matching process could then be directly and objectively scored in terms of success or failure, according as the numbers were correctly or incorrectly allotted to the personality descriptions.

*Statistical Methods*

Statistical analyses of the number of pictures correctly identified in the description and matching experiments are contained in Tables I and II. Each factor is considered separately. The total number of pictures placed by subjects in each category or Szondi factor is given, together with the proportion of these placings which agree with Szondi's diagnoses, expressed as percentage values. The probability that the total correct identifications for any given factor exceeds the number of correct identifications that would be expected by chance is calculated by working out the binomial distribution for each factor. The probability that a given Szondi factor will be correctly identified by chance is  $1/8$  or 12·5 per cent.; the probability that it will be incorrectly identified is  $7/8$  or 87·5 per cent. Therefore, the mean and variance of the binomial distribution can be calculated. The difference between chance and observed frequencies is divided by the standard deviation, which gives a figure for  $c$ —the normal deviate. The resulting  $P$  is obtained from tables given by Mather (1943).

Table III gives the analysis for each picture within each factor for both methods. Here again, the probability that a given picture will be correctly identified by chance is  $1/8$  or 12·5 per cent. The obtained percentage of correct identification for each picture is given, and is seen to be either above or below the percentage expected by chance; but the numbers involved are too small for  $P$  to be calculated. The table is sufficient to show the relative success with which different pictures were identified. In many cases, the percentage values show such clear differences that no further statistical refinements are needed.

## SUBJECTS

*(a) Description Method*

There were 24 subjects, including 15 men and 9 women. All the men were first or second year University students, mostly from the Science faculty; none were students of psychology. Eleven men wrote descriptions for all the 48 cards, four men completed from one to three sets only. Their average age was 19·9 years, with a S.D. of 2·01. Six of the women were first year Social Science students, three were hospital sisters with experience of psychiatric patients. Their average age was 29·4 years, with a S.D. of 14·36.

*(b) Matching Method*

There were 38 subjects, including 33 women and 5 men. Twenty-three subjects were members of a Women's Institute; the remaining two groups of 6 and 9 subjects respectively belonged to two different University Extension classes in psychology. The average age for all subjects was 47·3 years, with a S.D. of 11·39.

## RESULTS

In the entire collection of 48 pictures, a total of six pictures can be expected to be correctly identified by chance. The mean numbers of correct identifications in both experiments are considerably above this number. The figure obtained in the Description experiment is 14·3, and in the Matching test, 9·3. These figures exceed chance expectation at the 0·001 level of significance, if the  $t$  ratios are calculated from the percentages of total correct identifications to give the normal deviate. The normal deviate for results in the Description test is 20·7; for the Matching test it is 6·8.

But the results displayed in the tables show firstly, that some factors are more easily identified than others (Tables I and II); secondly, that pictures within each category are of grossly unequal stimulus value (Table III).

There is a considerable difference in results obtained, respectively, by the two different methods. Table I shows that the percentage of subject's identifications in the Description test exceeds chance expectation at the 0.001 level for six out of the eight factors; whereas similar percentages in the Matching test, given in Table II, are obtained for only three factors; and one of these, the Depressives, is not significantly identified in the Description test. There was an overlap of identifications in both methods in only 29 out of the 48 pictures.

However, in both experiments, the accurate identification of Homosexuals and Manics exceeds chance expectation, and the Epileptics are at or below the chance level. This is the only true measure of agreement in results obtained on factors as such by the two different methods.

Table III indicates that there is a considerable difference in both sets of results between the stimulus values of individual cards within each factor. Even where there is a high degree of correct identification of pictures within one factor, accurate diagnoses are always weighted on particular cards. For example, in the M factor, four cards are consistently identified with accuracy in both sets of results; whereas the two remaining cards are seldom identified—one of them, VI, 1, is never correctly identified by anybody.

TABLE I

*Analyses of Subject's Correct Identifications in the Description Test*

Szondi Categories	Total Number of Szondi Pictures in each category	Total Number of Pictures placed in each category	Percentage of Pictures correctly identified	c	P
Homo. . . . .	129	124	53.5	14.10	0.001
Sadis. . . . .	128	120	32.8	7.00	0.001
Epil. . . . .	129	100	17.1	1.59	0.20-0.10
Hyst. . . . .	129	127	22.5	3.40	0.001
Cat. . . . .	128	107	22.5	3.40	0.001
Par. . . . .	128	106	33.3	7.17	0.001
Dep. . . . .	129	103	16.2	1.28	0.20
Man. . . . .	129	156	51.9	13.60	0.001

TABLE II

*Analyses of Subject's Correct Identifications in the Matching Test*

Szondi Categories	Total Number of Szondi Pictures in each category	Total Number of Pictures placed in each category	Percentage of Pictures correctly identified	c	P
Homo. . . . .	225	234	42.7	13.70	0.001
Sadis. . . . .	225	302	12.4	0.04	0.95
Epil. . . . .	225	234	9.3	1.50	0.20-0.10
Hyst. . . . .	225	222	9.3	1.60	0.10
Cat. . . . .	225	196	11.1	0.64	0.50-0.60
Par. . . . .	225	217	13.8	0.59	0.50-0.60
Dep. . . . .	225	190	18.2	2.59	0.01
Man. . . . .	225	175	38.7	11.90	0.001

TABLE III  
*The Identification of Individual Pictures*

Szondi Categories	Percentage of Szondi Pictures correctly identified in each category	
	Description Method	Matching Method
<b>Homosexuals</b>		
I, 5 .. ..	31·8	13·2
II, 4 .. ..	54·5	52·6
III, 1 .. ..	59·1	39·5
IV, 8 .. ..	85·7	86·5
V, 7 .. ..	38·1	18·9
VI, 2 .. ..	52·4	48·6
<b>Sadists</b>		
I, 2 .. ..	36·4	2·6
II, 7 .. ..	81·8	5·3
III, 3 .. ..	9·5	5·3
IV, 6 .. ..	5·2	5·3
V, 5 .. ..	5·2	5·3
VI, 4 .. ..	57·1	51·4
<b>Epileptics</b>		
I, 6 .. ..	27·3	18·6
II, 3 .. ..	27·3	2·6
III, 2 .. ..	18·2	7·9
IV, 7 .. ..	9·5	23·7
V, 1 .. ..	19·0	2·7
VI, 8 .. ..	0·0	0·0
<b>Hysterics</b>		
I, 8 .. ..	5·6	7·9
II, 1 .. ..	40·9	2·6
III, 7 .. ..	18·2	0·0
IV, 2 .. ..	19·0	13·5
V, 3 .. ..	33·3	29·7
VI, 6 .. ..	19·0	2·7
<b>Catatronics</b>		
I, 1 .. ..	45·5	7·9
II, 8 .. ..	36·4	26·3
III, 5 .. ..	28·6	10·5
IV, 4 .. ..	9·5	13·5
V, 6 .. ..	14·3	8·1
VI, 3 .. ..	0·0	0·0
<b>Paranoiacs</b>		
I, 3 .. ..	22·7	15·8
II, 6 .. ..	36·4	15·8
III, 8 .. ..	19·0	0·0
IV, 1 .. ..	57·1	2·7
V, 4 .. ..	57·1	21·6
VI, 5 .. ..	9·5	27·0
<b>Depressives</b>		
I, 4 .. ..	31·8	39·5
II, 5 .. ..	22·7	5·3
III, 6 .. ..	18·2	35·1
IV, 3 .. ..	9·5	10·8
V, 2 .. ..	9·5	18·9
VI, 7 .. ..	5·2	0·0
<b>Manics</b>		
I, 7 .. ..	54·5	28·9
II, 2 .. ..	86·4	39·5
III, 4 .. ..	68·2	63·2
IV, 5 .. ..	9·5	8·1
V, 8 .. ..	90·5	91·9
VI, 1 .. ..	0·0	0·0

## DISCUSSION

Considering the total number of correct identifications obtained in both experiments, and the degree to which they exceed the number to be expected by chance, it is clear, at least, that the pictures are not meaningless stimuli for the subjects concerned. But some pictures are obviously more readily identified than others, and the whole question of scatter in the results is one that merits more detailed examination.

Firstly, it is somewhat difficult to account for the divergence in results between the Description and Matching experiments. The results show that those subjects who were allowed to write their own descriptions to the pictures quite freely, gave more correct identifications than the subjects who were confined to the checklist. On the face of it, a checklist should make identification easier; indeed, Rabin (1950a), remarks that it is probably much easier to diagnose with rather than without the aid of a checklist. The above results, however, indicate that the reverse is true for the subjects examined. One obvious reason for this may be that 23 out of the 38 subjects engaged in the Matching experiment were members of a Women's Institute. Some of these women found the novelty of the experimental situation disconcerting, and as a whole this group probably had less psychological insight than the students. Yet there was no appreciable difference in accuracy in matching shown by Institute members and accuracy attained by the 15 Extension class students. Another reason for the disparity between results may be the fact that some subjects reported that they found the checklist too constricting; that is, on occasion some of them wished to combine two or more personality descriptions in the identification of a particular picture, with a relative predominance of one description. Alternatively, they sometimes wanted to reject all descriptions and invent their own in the manner of the Description method. In the extensive experiments of Klopfer and Borstelmann (1950), this first difficulty was overcome by obtaining first and second choices in response to each picture. In this way, the two independent procedures they used arrived at the same designations of the stimulus pictures in 41 out of the 48 pictures—a considerably greater overlap than that reported above.

Another possible advantage held by the Description method is that the technique of uncontrolled association to pictures corresponds closely to Szondi's own method of picture selection. Bearing in mind the Szondi-Deri rationale, a certain liberality therefore was possible in scoring results obtained from the Description method which could not be exercised with the Matching method without making the checklist personality descriptions much too complicated for subjects to absorb readily. Scoring was "liberal" in the sense that the various "sublimations" of mental disorders described by Deri (1949), as surgeons, spies, wrestlers, etc., were scored as correct identifications where they occurred in appropriate categories. The models used in this scoring are contained in chapter III of Deri's book.

Despite this disagreement in the actual significance in the general results obtained from the two experiments, it is firmly established that the Homosexuals and Manics are more easily identified than any of the other factors in both experiments. This confirms the findings of Klopfer and Borstelmann (1950), and of Rabin (1950a and b). Why these two factors in particular are more readily identified than the others is a matter for some speculation. It is possible that superficial facial clues determine responses rather than the genetic determinants suggested by Szondi or the latent tendencies hypothesized by Deri. If this is true, then any smiling face might be classed as Manic, and this would explain why card IV, 5 is so seldom identified, and why card VI, 1 is never



identified at all. It is more difficult, however, to single out definite facial characteristics by which the homosexuals might be easily identified. Yet if all the six H pictures are compared one with another, they all appear to look younger and less mature than most of the remaining pictures; the lips especially suggesting youth and immaturity. There is a notable exception with card I, 5, of an older man, with an entirely different cast of features than the other homosexuals, and this picture is much less frequently identified than the remainder in both experiments (see Table III). It is probable that facial clues are of some considerable importance in identification, because experiments conducted by Wallen (1951), on facial tension shown in the Szondi pictures, led him to claim that "groupings based on facial characteristics are more meaningful than those based on the diagnoses of the persons photographed". This might also suggest why the depressives were also identified above chance expectation in the Matching experiment. The drooping mouth might be sufficient to classify at least three out of the six pictures in the D category. The reason why the Depressives were not similarly identified in the Description experiment may be because Deri's character "sublimations" are often far removed from what would be expected from an unhappy facial expression.

This inequality of stimulus value of pictures within the D category is characteristic of all the other Szondi categories (Table III). Inconsistent identifications of pictures within categories are also reported by Dudek and Patterson (1952), using the Matching method. Davis and Raimy (1952), working on 24 cards only, also discovered stimulus inequalities; and the conclusion of Guertin (1951), who used factor analysis on the pictures, is that "there is as much difference in meaning between pictures within the same diagnostic category as there is between pictures from different categories". This finding, as the author points out, is quite at variance with Szondi's assumption that pictures within a diagnostic category have a similar meaning or diagnostic significance. The obvious conclusion to be drawn from these results is that a complete re-organization of the test is needed, where each picture has a known stimulus value of precise statistical significance, standardized for a particular population.

In this event, it is recommended that the Epileptic factor should be excluded from the test altogether. There is now a considerable weight of experimental evidence to support this, not only in the above findings (Tables I and II), but in the results Rabin obtained from students (1950a), and those obtained from a larger population by Klopfer and Borstelmann. All these results show that the E factor is the one least often identified with accuracy. Epilepsy, in particular, raises the further question of the etiologies of the disorders represented by the different Szondi factors. In view of the extreme variety of the causes of epilepsy, it is, even a priori, highly unlikely that there is a personality stereotype for "epilepsy" as such; and this view is supported by the experience of clinicians and the findings of projection tests such as the Rorschach. Even with homosexuals, who are much more successfully identified, it is common knowledge that various environmental and sociological factors may be responsible for the homosexual behaviour which designates a person as "homosexual"—to say nothing of the question of endocrine imbalance. The resulting personality type will depend on the relative predominance of these factors. Yet Szondi has nothing to say on this question of etiology.

At this stage it might be argued that in spite of the dubious nature of the theoretical rationale of the Szondi test, it is nevertheless a clinical tool which can be used in research merely because it seems to "work" in clinical practice. But

a preliminary review of the literature strongly suggests that even the clinical value of the test is by no means established.

Szondi literature at present reveals seven studies which support the test in its clinical aspects. These are by Calabresi (1948), David (1949), Holt (1950), Kaldegg (1951), Stumper (1951), Kobler (1952), and by Mignot and Gabel (1952). Eight studies, on the other hand, did not reveal the test to be reliable from many clinical points of view. These studies are by Ancelin *et al.* (1949), Giuliani (1949), Chiodi *et al.* (1950), Fosberg (1950), Paine (1950), Wiegersma (1950), Cole (1951), and by Mussen and Krauss (1952).

This review is somewhat more favourable to the test than the one presented by Guertin and McMahan (1951), who found only two investigations which seemed to support the test clinically. Yet six of the above seven pro-Szondi studies were conducted on four or less than four subjects; the remaining study was only a short preliminary report (by David, 1949). Here, results based on Szondi's method of structural analysis agreed with staff diagnosis in 88 per cent. cases of 50 paranoid schizophrenics. Yet as Guertin and McMahan point out, "the report is so brief that the reader is left in doubt as to what constitutes agreement". On the other hand, studies raising critical doubts about the value of the test were all conducted much more extensively on larger populations, never on less than twenty subjects.

On balance, therefore, it is very doubtful whether the Szondi test as it stands at present is of sound clinical value. This makes a demonstrable theoretical rationale doubly necessary if the test continues to be used in clinical practice. The methods described of obtaining verbal reactions to the pictures and assessing their statistical significance in relation to the Szondi factors is a fruitful approach to the problem of validation. Enough has been demonstrated in the Tables and elsewhere to show that it is doubtful whether many of the individual pictures, and at least one factor—the Epileptic—should be retained. Future use of the Szondi test should be guided by the relative degrees of reliability of the various categories and cards discovered in this investigation and the other researches mentioned. In view of the considerable amount of agreement reached by different workers, a revised scale could be quite easily worked out on the basis of their findings. This would be of greater value to clinical psychologists than the test as it stands at present, with its inherent inconsistencies in stimulus values.

#### SUMMARY

An attempt was made to investigate the theoretical basis of the Szondi test by obtaining, experimentally, identifications of the Szondi pictures. The Description and Matching methods used correspond to those adopted by American investigators, but were employed for the first time on groups of English subjects.

The results show that although, in general, the pictures are not meaningless stimuli, certain factors and particular pictures are more readily identified than others. Homosexuals and Manics are the most frequently identified factors in both tests, the Epileptics being least frequently diagnosed with accuracy. These findings largely confirm the results published by American investigators, although significant differences were found in results obtained respectively from the two different methods.

It is suggested further that facial clues may be important in the determination of responses. The inequality of stimulus value of particular cards indicates that a complete reorganization of the test is needed. This might now be achieved by a synthesis of the results of several experimental studies available on Szondi theory.

#### ACKNOWLEDGMENT

I should like to thank Professor F. L. Golla for his stimulating interest in this study.

#### REFERENCES

- ANCELIN, A. E., DUCHENE, H., and SCHÜTZENBERGER, M., *Rev. Psicol. gen. apl.*, Madrid, 1949, 4, 437 (abstract).



- CALABRESI, R., *Rorschach. Res. Exch.*, 1948, **12**, 238 (abstract).
- CHIODI, V., GILLI, R., and GIULIANI, O., *Riv. Psicol.*, 1950, **46**, 181 (abstract).
- COLE, D., *J. clin. Psychol.*, 1951, **7**, 383.
- DAVID, H. P., *Szondi Newsletter* (mimeographed by University of Nebraska), 1949, **1** (in Guertin and McMahan).
- DAVIS, N. E., and RAIMY, V. C., *J. clin. Psychol.*, 1952, **8**, 155.
- DERI, S., *Introduction to the Szondi Test: Theory and Practice*, 1949. New York: Grune and Stratton.
- DUDEK, F. J., and PATTERSON, H. O., *J. consult. Psychol.*, 1952, **16**, 389.
- FOSBERG, I. A., *Amer. Psychol.*, 1950, **5**, 326.
- Idem*, *J. consult. Psychol.*, 1951, **15**, 39.
- GIULIANI, O., *Riv. Psicol.*, 1949, **45**, 143 (abstract).
- GUERTIN, W. H., *J. clin. Psychol.*, 1951, **7**, 232.
- Idem* and MCMAHAN, H. G., *Amer. J. Psychiat.*, 1951, **108**, 180.
- HOLT, R. R., *J. proj. Tech.*, 1950, **14**, 435 (abstract).
- KALDEGG, A., *J. Ment. Sci.*, 1951, **97**, 555.
- KLOPFER, W. G., and BORSTELMANN, L. J., *J. Personality*, 1950, **19**, 172.
- KOBLER, A. C., *Encéphale*, 1952, **8**, 125.
- MATHER, K., *Statistical Analysis in Biology*, 1943. London: Methuen.
- MIGNOT, H., and GABEL, J., *Rev. psychol. appl.*, 1952, **2**, 39 (abstract).
- MUSSEN, P. H., and KRAUSS, S. R., *J. abnorm. soc. Psychol.*, 1952, **47**, 399.
- PAINE, H. E., *Amer. Psychol.*, 1950, **5**, 481.
- RABIN, A. I., *J. abnorm. soc. Psychol.*, 1950a, **45**, 392.
- Idem*, *J. consult. Psychol.*, 1950b, **14**, 400.
- STUMPER, E., *Ann. med.-psychol.*, 1951, **109**, (1:3), 280 (abstract).
- SZONDI, L., *Experimental diagnostics of drives*, 1952. New York: Grune and Stratton.
- WALLEN, R., *J. consult. Psychol.*, 1951, **15**, 210.
- WIEGERSMA, S., *Ned. Tijdschr. Psychol.*, 1950, **5**, 108 (abstract).